BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e181968
doi: 10.3897/biss.9.181968

OPEN ACCESS

Conference Abstract

AI-Assisted Exploration, Curation, and Extension of Biodiversity Data Using iChatBio

Michael J. Elliott‡, Manuel Luciano‡, Jose Fortes‡

‡ University of Florida, Gainesville, Florida, United States of America

Corresponding author: Michael J. Elliott (mielliott@ufl.edu)

Received: 10 Dec 2025 | Published: 11 Dec 2025

Citation: Elliott MJ, Luciano M, Fortes J (2025) AI-Assisted Exploration, Curation, and Extension of Biodiversity Data Using iChatBio. Biodiversity Information Science and Standards 9: e181968.
https://doi.org/10.3897/biss.9.181968

Abstract

The ability of large language models (LLMs) to interact with their environment through function-calling mechanisms has enabled the development of *agentic systems* - artificial intelligence (AI) systems that can:

1. understand natural language requests,
2. create an action plan to address the response (e.g., retrieve data from an online source),
3. execute the action plan through the use of *tools*, and then
4. interpret the outcomes of their actions to generate natural language response.

As research tools, agentic systems have the potential to transform how researchers interact with scientific data. They allow users to interact with the system through *requests* rather than *instructions* (though users may still issue instructions if desired), letting agents autonomously plan and execute the labor-intensive tasks of finding, retrieving, processing, and compiling information across one or more information sources.

General-purpose agentic systems struggle to make effective use of scientific data and retrieval tools due to lack of domain expertise, and also produce information that is difficult to verify. To address these issues, we have created iChatBio, an agentic system specifically designed to a) incorporate expert biodiversity knowledge into its decision-making processes and b) provide traceability of information back to sources. This is

© Elliott M et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

achieved through a multi-agent design in which one general-purpose *chat agent* breaks down complex user requests into specific information needs which are fulfilled by *expert agents*, which are specialized to find or process certain types of information. iChatBio currently connects with expert agents that access data served by iDigBio*[1], GBIF*[2], iNaturalist, BHL*[3], OBIS*[4], and ALA*[5], with more expert agents being added regularly. iChatBio also connects with several expert agents that provide generic data processing and visualization functions. This synergistic combination of data retrieval experts and data processing experts enables new, emergent data use cases beyond the scope of any individual data provider. This presentation discusses the following use cases:

- **Exploration:** includes finding, retrieving, summarizing, and comparing biodiversity data from diverse sources
- **Curation:** includes compiling existing records into new thematic datasets, fixing spelling and formatting errors, combining duplicate records, and inferring missing field values
- **Extension:** includes linking specimen records to related records, genetic data, literature, and more, to form extended specimen records

iChatBio makes integrated data services more easily accessible through its LLM-powered natural language interface. For example, the use of multilingual LLMs like OpenAI's GPT-4.1 allows users to interact with biodiversity portals and the data they serve in the user's native language (so long as the LLM is proficient in the language). Additionally, although human-friendly data portals are often designed to support specific use cases, data providers also often host machine-friendly Application Programming Interfaces (APIs) which support many more use cases through highly configurable functions. Such APIs typically require significant effort and expertise on the user's part to use effectively; iChatBio's expert agents, however, allow users to make use of such APIs without any experience. Expert agents can even educate users about APIs by demonstrating how to construct API calls for particular use cases.

iChatBio is an evolution of the agentic system presented in Elliott et al. 2024, improving upon the original system in both system design and breadth of integrations with biodiversity data services. Whereas the original system design was *monolithic* (all agents existed as part of a single software application), iChatBio is a *distributed system* (agents run as independent web servers) in which expert agents communicate over the Internet using the Agent2Agent (A2A) protocol (Surapaneni et al. 2025). This distributed design enables community-driven agent development, where expert agents can be independently managed and improved by community members. Moreover, A2A compliance allows expert agents developed for iChatBio to potentially participate in other agentic systems, and likewise existing A2A-supporting agents can potentially connect to iChatBio. To facilitate community contributions of expert agents, we have assembled an iChatBio Software Development Kit (iChatBio SDK), available in the Python Package Index.

Keywords

LLM, agent, API, integration, chat

Presenting author

Michael Elliott

Presented at

Living Data 2025

Acknowledgements

The research reported in this work was funded in part by grants from the National Science Foundation (DBI 2027654) and the AT&T Foundation.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Elliott M, Luciano M, Fortes J (2024) Integrating Large Language Models and the iDigBio Portal for Conversational Data Exploration and Retrieval. Biodiversity Information Science and Standards 8 https://doi.org/10.3897/biss.8.142696
- Surapaneni R, Jha M, Vakoc M, Segal T (2025) Announcing the Agent2Agent Protocol (A2A). Google for Developers URL: https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/

Endnotes

*1 Integrated Digitized Biocollections
*2 Global Biodiversity Information Facility
*3 Biodiversity Heritage Library
*4 Ocean Biodiversity Information System
*5 Atlas of Living Australia